

Advanced 3D Chips Show Promise *But Conventional 3Dlabs' Glint Outshines More Radical Designs*

by Peter N. Glaskowsky



Most recently announced 3D accelerators, including those covered in our previous article (see MPR 6/2/97, p. 16), are single-chip solutions that deal only with setup and rendering. They implement conventional 3D pipelines designed to re-render every polygon for every frame.

This approach is easy to understand and implement, but it is too restrictive for today's high-end applications as well as tomorrow's mainstream 3D. Multichip implementations, programmability, and revolutionary new rendering architectures are needed to meet these challenges.

Some vendors are following an evolutionary path to higher performance, extending their existing architectures with the aid of denser, faster fab processes. Others, including Rendition, RSSI, and VideoLogic, have already discarded the conventional 3D pipeline as obsolete and are pursuing more radical approaches, hoping to establish themselves as the leaders of the 3D market.

It's unlikely that any of these more aggressive designs will displace the conventional solutions in the next year or two, but it's clear that today's conservative architectures won't survive long into the next millennium. For 1997, we give the nod to 3Dlabs' Glint family as having the best performance of all announced 3D accelerators despite mostly incremental improvements over previous Glint chips. The combination of the Glint MX rendering engine and the Glint Gamma geometry processor enables sustained rendering rates in excess of 2 Mtriangles/s and 55 Mpixels/s, well above competing solutions for personal computers.

Opening Up 3D Bottlenecks

Geometry processing and memory bandwidth represent the two most pressing problems for 3D performance. Host processors alone cannot provide enough vertex data to keep even an inexpensive rendering chip busy. Likewise, without fast, wide local memory, a 3D chip cannot perform texture mapping with high quality and a fast update rate.

There are two basic responses to these challenges. The more common is to apply more transistors and package pins. This is certainly a feasible solution for the time being, and it will always remain a factor in successful 3D designs, but in the long run, finesse will win out over brute force. Designers will find smarter ways to achieve the same visual effect with fewer calculations and less memory bandwidth.

3Dlabs' Glint Provides Scalable Performance

A pioneer in PC 3D, 3Dlabs is now on its third generation of CAD-oriented 3D chips. The Glint 300SX introduced OpenGL hardware acceleration to Windows NT systems in 1994, with the 500TX adding texture-mapping support in late 1995. The latest Glint rendering chip, the MX, greatly improves texturing throughput and overall performance.

If the Glint MX were 3Dlabs' only contribution to high-end 3D, it would be reasonably successful—but the MX is only part of the company's strategy. The other key component of the Glint family is the new Glint Gamma geometry processor, capable of offloading all 3D geometry, lighting, and setup calculations from the host processor. Used in conjunction with the MX, as Figure 1 shows, Gamma performs at a rate of 1 GFLOPS and acts as an AGP-to-PCI-66 bridge for the MX, enabling single-slot solutions with one Gamma and up to eight MX chips.

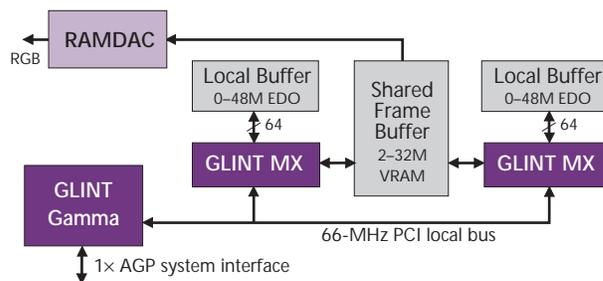


Figure 1. The 3Dlabs Glint Gamma and Glint MX support multi-chip implementations as well as unusually large local memory arrays, supporting high-resolution displays with extremely high sustained throughput.

Even with setup processing handled by the 3D chip, a 266-MHz Pentium II can generate just 500,000 lit, textured triangles per second for OpenGL rendering, using 100% of the processor. Gamma can perform geometry and setup calculations for 3.3 million triangles per second while using only a small fraction of the host's processing power, greatly increasing application performance and visual quality. The higher triangle rate translates directly to better frame rates.

A few other 3D accelerators, such as TriTech's Pyramid3D (see MPR 11/18/96, p. 5), also include geometry accelerators, but Gamma is the first to bring this level of performance to the personal computer market.

With setup processing handled by Gamma, the MX does not need its own on-chip setup processor, a common feature in other recent 3D chips; instead, all of the MX is dedicated to 3D rendering, as Figure 2 shows. Each MX can render about one million triangles per second and perform texture mapping at 33 Mpixels/s with bilinear filtering. Trilinear filtering is also supported, requiring eight texel reads per pixel, compared with four for bilinear filtering; in this mode, throughput drops to 16.5 Mpixels/s, still faster than trilinear filtering on most competing chips.

The performance of a single MX is very good, but the part really shines when used in pairs or quad arrays. In such arrangements, successive scan lines on the display are assigned to each MX in order. Polygons that span multiple scan lines are then rendered cooperatively by all the affected MX chips, distributing the load with reasonable efficiency. As a result, two Glint MXs in conjunction with a single Gamma can render 2 Mtriangles/s and 55 Mpixels/s.

Another approach would have been to divide the screen into two or more regions, assigning each region to a single MX chip, but 3Dlabs felt this option would have led to less-predictable performance.

High-performance texturing is a relatively recent addition to the Glint family. While the 500TX included a texturing engine, that chip did not support MIP mapping, and its

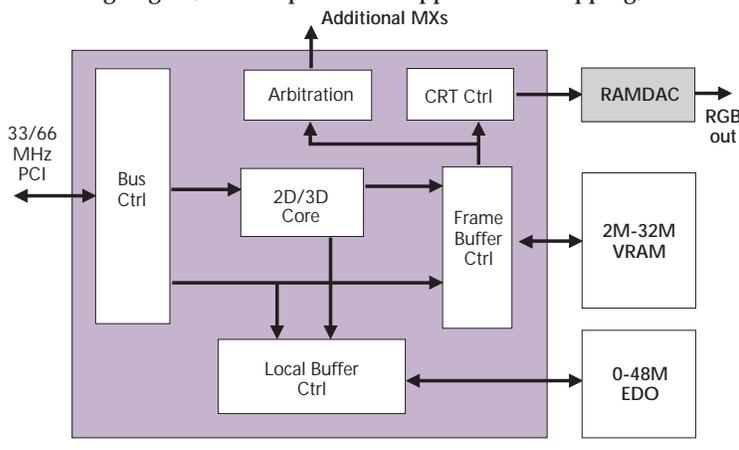


Figure 2. The 3Dlabs Glint MX is a fairly conventional 3D-rendering chip intended for use with the Glint Gamma geometry processor. As a result, the MX includes no setup processor and is entirely devoted to high-performance, high-resolution rendering.

performance was just a fraction of the MX's—4.5 Mpixels/s compared with 33 Mpixels/s.

This increased attention to texturing performance represents two interesting trends in PC 3D. First, CAD operators are becoming more interested in the ability to apply textures to 3D models under development. Second, the performance penalty for texturing, once substantial, has been greatly reduced, due to architectural innovations in modern rendering engines. For example, the MX has a separate texture memory interface, allowing it to read and filter texture data as quickly as it can render and shade the pixels.

Support for Multiple Chips Provides Scalability

3Dlabs' focus on these high-end configurations also influenced its choice of system interfaces for Gamma and the MX. The high throughput of AGP is necessary for scenes with high polygon counts, but AGP does not support multiple devices. To get around this limit in multiple-MX configurations, the Gamma geometry processor is equipped with both AGP and PCI interfaces. The chip receives vertex data from the host over the AGP interface; it then uses the subsidiary PCI bus to communicate with one to eight MX chips or other PCI graphics devices such as video digitizers.

Boards with four MX devices can run the local PCI bus at 66 MHz, but because of PCI's electrical loading limits, eight-chip configurations must run the local bus at 33 MHz. This reduces the peak polygon throughput of the local bus, which may be acceptable for those few applications that depend on high fill rates and have low polygon counts, but we expect to see the highest performance for typical applications from the two- and four-MX configurations.

Another indicator of the MX's high-end orientation is its support for ultrahigh resolutions, up to $2,048 \times 2,048$ pixels. In particular, the MX is designed to support HDTV-class $1,920 \times 1,200$ -pixel resolutions on monitors such as Sony's wide-screen GDM-W900. The MX can be used with up to 32M of frame-buffer memory, which is more than adequate for HDTV-resolution true-color 3D rendering. Multiple MX chips share a single frame buffer. The MX's frame buffer is implemented with VRAM, an older memory technology than the SGRAM found on most recent 3D chips. Because VRAM is dual-ported, screen-refresh traffic is removed from the frame-buffer interface, saving as much as 500 Mbytes/s for high-resolution displays.

In addition to the 32M of frame buffer, the MX can handle another 48M of EDO DRAM for its local buffer, where Z, stencil, and texture data are stored. Since this data is not sent to the monitor, there is no need for VRAM's serial port, and EDO DRAM is considerably less expensive. Each MX in a multichip subsystem has its own local buffer, so in theory a vendor could assemble an MX-based card with 32M of frame buffer plus 384M of local buffer.

Gamma/MX graphics cards will not be cheap. A typical dual-MX board with 32M of frame buffer and 32M of local buffer will have a materials cost of more than \$900, far higher

than the other products reviewed in this series. Due primarily to Gamma's geometry acceleration, the resulting product will be about three times faster than any competing graphics card, however, more than justifying the price premium in high-end applications like 3D CAD.

Rendition Revises VÉRITÉ Family

Rendition is one of the more innovative companies in the PC 3D graphics market. The company's V1000 chip (see MPR 5/6/96, p. 1) incorporates a RISC CPU core as well as a 3D-rendering engine; it was the first mainstream chip to provide a texture cache, and it offered better texturing performance than competing single-chip products.

The new V2200, described in general terms at PC Tech Forum, will be a fairly straightforward speedup of the V1000. While the V2200 will include an upgraded version of the V1000's RISC core, Rendition does not emphasize the part's programmability. Instead, the V2200 will perform most operations in fixed-function logic blocks.

Rendition is designing the V2200 with a 1× AGP interface, feeling that 2× AGP is not yet necessary due to restricted main-memory bandwidth. Subsequent members of the V2000 family are likely to include 2× AGP support once PCs with 100-MHz SDRAM become available.

The V2200 will support a local 64-bit SGRAM array up to 16M in size. This will allow the V2200 to manage a true-color 1,024 × 768-pixel frame buffer with Z buffering while preserving over 8M for local texture storage.

RSSI PixelSquirt Dispenses With Frame Buffer

The Tex SPC1516 from RSSI is certainly the most unusual of today's 3D chips for the PC. It completely dispenses with the usual polygon-at-a-time scheme used by other mainstream chips, using instead 256 parallel processing elements in four parallel pipelines to render up to one scan line at a time from a list of presorted polygons prepared by the system's host processor.

Tex is the second in a series of chips based on RSSI's PixelSquirt architecture. The chip follows the PIX SPC1515 (see MPR 5/6/96, p. 5), which found design wins at a few OEMs, including Apple.

Tex also dispenses with a frame buffer and even a CRT controller by working with a separate graphics chip that manages the display. Like 3Dlabs' Gamma, Tex can act as a limited AGP-PCI bridge, creating what RSSI calls a Pixel Pipe interface to a local graphics chip. This enables single-board AGP products, as Figure 3 shows.

Like its predecessor, Tex can also reside on its own card or on the motherboard, sending its display data to a graphics chip or card located elsewhere in the system. The latter configuration is likely to be unacceptably slow, however, since the system's PCI bus would be required to carry all vertex data plus the rendered 3D pixel data; RSSI recognizes this bottleneck and does not recommend this configuration. Some higher-resolution display modes would not work at all

because of inadequate bandwidth in the shared PCI system environment.

Because Tex does not use a local frame buffer, it can operate with little or no local memory. The chip has an internal texture cache and also supports an optional 8M texture cache in off-chip SGRAM. Additional texture data may be stored in host memory and accessed over AGP, or in frame-buffer memory when Tex is used with a local graphics chip.

Tex's integrated setup engine and multiprocessor renderer allow very high peak performance. Most texture reads, as well as all Z- and color-buffer information, are satisfied from on-chip memory during rendering. This reduces the typical memory-bandwidth bottleneck, which still remains the primary obstacle to better performance from this design.

RSSI claims a sustained throughput of 530 Ktriangles/s and 40 Mpixels/s with trilinear MIP-mapped textures, but the company has not yet released 3D WinBench numbers. The chip's peak pixel-fill rate with bilinear filtering is in excess of 150 Mpixels/s—impressive enough by itself, but Tex achieves this level of performance regardless of the amount of triangle overlap in the rendered scene. In most 3D chips, triangle overlap—called depth complexity—results in wasted rendering effort. If an average of three triangles cover each pixel on the screen, the effective fill rate of conventional 3D chips is cut by a factor of three. Tex renders such scenes with no loss of performance regardless of depth complexity, enabling 1,024 × 768-pixel true-color rendering at about 53 frames per second, which the company claims is faster than any other mainstream part.

There are some caveats to this achievement. Depth complexity is actually limited to 255 triangles per average pixel on a single scan line; if this number is exceeded, Tex will split the scan line into multiple segments and render the line properly, but with some loss in performance. Tex's video driver software is required to make this decision during the rendering process. On the other hand, scenes in today's software rarely have average depth complexities

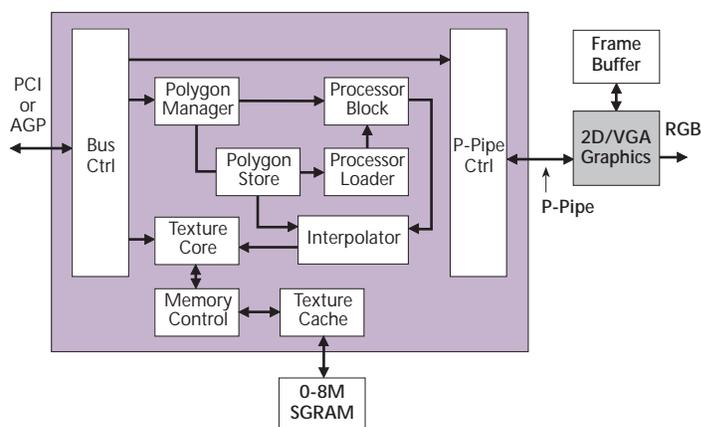


Figure 3. The RSSI Tex performs 3D rendering on a per-pixel basis using 256 processing elements and on-chip memory. Rendered pixels are sent to a separate 2D graphics chip to be displayed.

Price and Availability

Table 1 shows pricing and availability information. Contact 3Dlabs (San Jose, Calif.) at 408.436.3455 or visit the company's Web site at www.3dlabs.com. Contact Rendition (Mountain View, Calif.) at 415.335.5900 or www.rendition.com. RSSI (San Jose, Calif.) can be reached at 408.435.5565 or on the Web at www.simsys.com. For information on the VideoLogic/NEC PowerVR, contact NEC (Mountain View, Calif.) at 415.965.6000 or access the Web at www.PowerVR.com.

much above 1.5. Typical 3D games have backgrounds plus a few foreground elements that rarely overlap. Conventional 3D chips render these scenes adequately, but over time we expect to see depth complexity increasing as 3D software vendors produce titles with more complicated scenes, giving the PixelSquirt architecture an increasing edge.

Tex has another unique distinction: it is the first PC 3D chip to offer a form of anisotropic texture filtering, producing higher visual quality for textures applied to objects at a sharp angle to the viewpoint, such as building facades in a driving simulation. RSSI's technique is like that used by Microsoft in the Talisman architecture, using a different texel-sampling pattern but achieving similar results. We expect anisotropic filtering to be one of the key differentiating features among 3D chips in 1998, but for now, there is no support for anisotropic filtering in Direct3D, so this feature in Tex is likely to go largely unused unless software vendors choose to support it directly in applications.

VideoLogic and NEC Update PowerVR

VideoLogic's PowerVR architecture, codeveloped and marketed by NEC, has developed a reputation for high performance and quirky operation since its debut in early 1996 (see MPR 3/5/96, p. 16). In contrast to the frame-oriented rendering of Glint and Vérité and the pixel-oriented

rendering of PixelSquirt, PowerVR renders the screen one block at a time before moving on to the next block. The rendering takes place entirely in on-chip memory, gaining the same benefit from high on-chip bandwidth as PixelSquirt.

The latest member of the PowerVR family is the PCX2, a pin-compatible upgrade of the PCX1, which has been moderately successful in the PC games market over the past year. Both chips evolved from a scalable multichip implementation consisting of an image synthesis processor (ISP) and a texture and shading processor (TSP). For high-end products like arcade video games, multiple ISPs and TSPs could be combined. The PCX1 and PCX2 chips, intended for the mainstream PC market, have the equivalent of one ISP and one TSP in one device.

The PCX2's block sizes can be varied under program control according to scene complexity. Block sizes of 32 × 32, 32 × 64, and 64 × 64 pixels are supported. The actual rendering is done 32 pixels at a time by 32 on-chip parallel processing elements. Due to the inherently parallel nature of 3D rendering, we expect to see more vendors taking this approach in the future, whether in single-chip solutions like Tex and PCX2 or in multichip implementations like those possible with Glint MX.

The block-oriented rendering process eliminates the need for a local frame buffer, and like RSSI's Tex, the PCX2 needs only a moderate amount of local memory for a texture cache. From 1M to 4M of local memory is supported, with host memory used to store additional texture maps when needed. A fast on-chip 4K texture cache further accelerates texture mapping, and VideoLogic claims a respectable 40 Mpixel/s pixel-fill rate for bilinear-filtered textures. The PCX2 includes an on-chip setup processor, and the chip's polygon rate is given as 1.5 Mtriangles/s peak, sustaining slightly over 500 Ktriangles/s when used with a 200-MHz Pentium Pro CPU. Like RSSI, VideoLogic has not released 3D Winbench numbers for PCX2, but we expect the part to offer competitive performance.

Another characteristic the PCX2 shares with Tex is the lack of support for VGA graphics. Like Tex, the PCX2 must be used in combination with a separate 2D graphics chip. Unlike RSSI, however, VideoLogic has provided no pass-through PCI bus, preventing the design of complete 2D/3D products on a single AGP card. All rendered 3D graphics must be sent over the system's PCI bus to a separate 2D chip on a card or on the motherboard. This is likely to be a serious competitive disadvantage for PowerVR in the PC market, though not for arcade games. Few end users are prepared to deal with the inconvenience and added cost of multiple-card graphics subsystems.

	3Dlabs Gamma/MX	Rendition Vérité V2200	RSSI Tex	VideoLogic/NEC PCX2
Bus interface	AGP 1x	AGP 1x	66-MHz PCI	66-MHz PCI
Fastest memory type	VRAM+DRAM	SGRAM	SGRAM	SDRAM
Memory width	64+64	64	64	64
Memory clock rate	66 MHz	n/a	66 MHz	66 MHz
Maximum local RAM	32M+48M	16M	4M	4M
Texture cache	n/a	n/a	4K	4K
Setup engine	Full geometry	Yes	Yes	Yes
Peak triangle rate	3.3M	n/a	1.2M	1.5M
Peak pixel rate	33M per MX	n/a	150M	40M
RAMDAC	No	Yes	No	No
Availability	2H97	n/a	4Q97	Now
List price (1,000s)	\$525/set	n/a	\$35	\$35

Table 1. The chips described here take four different approaches to 3D rendering. The 3Dlabs Glint solution is fastest and most expensive, but the RSSI and VideoLogic products better represent the future of PC 3D. n/a: not available (Source: vendors)

Conventional 3D Still the Best—But Not for Long

While it seems inevitable that conventional single-pipeline, frame-oriented rendering will eventually become obsolete, today's best solutions still take this approach. The more revolutionary architectures, like PixelSquirt and PowerVR, show a great deal of promise but require complicated chip designs. In some cases, chip designers have chosen to provide the necessary complexity at the expense of valuable features like VGA compatibility.

3Dlabs, with more experience in PC 3D than any of its competitors, has produced the best high-end solution (Glint) as well as a strong mainstream product (Permedia). The company's attention to the needs of its broad customer base is clearly demonstrated by Permedia's high level of integration and Glint's multichip, high-performance design. The Glint Gamma solves the most critical problem for 3D CAD users, the geometry-processing bottleneck that slows rendering of complex 3D objects and assemblies. While a single Glint MX is not the fastest of all 3D-rendering engines, 3Dlabs allows multiple MX chips to be used in parallel to achieve best-in-class performance.

Rendition's on-chip RISC processor core, while not the centerpiece of the Vérité V2200's design, will provide valuable flexibility for video-related operations. As 3D chips are called upon to support higher-level functions such as geometry acceleration and display-list management, we expect to see Rendition and other vendors adding more programmability to manage these tasks.

RSSI and VideoLogic, with the most revolutionary 3D architectures we have covered here, have found interesting ways to solve some problems that are not yet crippling to their competitors but which are likely to become so within a few years. On-chip rendering solves the bottleneck in local-memory bandwidth, allowing higher quality and faster pixel-fill rates than conventional solutions. Today's Tex and PCX2 chips are not the best mainstream 3D solutions, but they point the way to excellent products to come.

Even more radical 3D architectures will appear during the next year. Trident plans to build a single-chip implementation of Microsoft's Talisman architecture (see MPR 8/26/96, p. 5), and other 3D chip companies, including sales-volume leader S3, have also announced Talisman plans. Talisman, like PowerVR, uses block-oriented on-chip rendering to reduce memory bandwidth. Like 3Dlabs' Gamma, Talisman is designed to work with a geometry accelerator to offload time-consuming floating-point coordinate transformations from the host processor. Uniquely, Talisman also eliminates some types of spatial and temporal redundancy from the 3D-rendering process, achieving even higher performance while reducing processing and memory-bandwidth demands even further.

Today's PC graphics market is quite interesting, but the coming year will be even more exciting as conventional architectures vie with radically different approaches to the goal of increasing visual realism for 3D. 

PC 3D Attracts a Throng of Companies

The PC 3D graphics market is in a constant state of flux. Just since our last issue, two vendors have entered the fray (NeoMagic and SiS) and two have left (Brooktree and S-MOS).

Following in the footsteps of European mapmakers during World War II, we are trying to keep track of the rapidly changing PC 3D landscape. The following is a summary of all the companies that have announced 3D chips for PCs, or plans to develop such products, as of our press deadline:

3Dfx (Voodoo Graphics); 3Dlabs (Permedia, Glint); Alliance (ProMotion); Artist (3GA); ATI (Rage); Avance Logic (ALG27000); Chromatic (Mpact); Cirrus (Laguna 3D); Dynamic Pictures (Oxygen); IGS (CyberPro 3000); Intel (740); Intergraph (Realizm); IXMICRO (TwinTurbo 128-3D); Matrox (Millennium, Mystique); Microsoft (Talisman); NeoMagic (MagicGraph 128XD); Number Nine (Ticket To Ride); Nvidia/SGS-Thomson (RIVA 128); Oak (Eon); Philips (Big Cats, TriMedia); Real3D (R3D); Rendition (Vérité); RSSI (PixelSquirt); S3 (Virge); Sigma Designs (RealMagic 3D); Silicon Reality (TAZ); SiS (SiS6326); Trident (3DImage); TriTech (Pyramid3D); Tseng Labs (ET6300); VideoLogic/NEC (PowerVR); VSIS (3D Pro); Yamaha (YGV612).

The majority of these 33 companies are developing conventional 3D-rendering chips with few if any differentiating features. As 3D chips become more sophisticated, the cost of developing new 3D products is rapidly approaching the cost of developing a high-end RISC CPU—\$10 million to \$30 million. The total available revenue from 3D chips is limited by the number of PCs sold (about 100 million in 1998) and the average cost of a graphics chip (about \$20).

A 20% margin leaves about \$400 million per year for 3D chip R&D, more than enough for everyone—but these funds are not evenly distributed. Major vendors like S3 and ATI can afford to continue their R&D efforts, but smaller companies will fall behind.

The Microsoft-led Talisman initiative offers renewed hope to the 3D companies that are already falling behind. Talisman represents the state of the art in 3D acceleration technology and is available for very moderate licensing fees. Talisman merely shifts the focus of the competition from architectural innovation to fab efficiency, however, an even tougher battle for many of these companies to win, given that most are fabless design houses.

We doubt that more than 10 of these 33 vendors will survive past the end of the decade. Deep pockets and sharp designers are equally important for companies wishing to beat the odds.